BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

**LIEFF CABRASER HEIMANN &
BERNSTEIN, LLP**
Rachel Geman (*pro hac vice*)
250 Hudson Street, 8th Floor
New York, New York 10013-1413
(212) 355-9500
rgeman@lchb.com

**CAFFERTY CLOBES MERIWETHER &
SPRENGEL LLP**
Bryan L. Clobes (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312) 782-4880
bclobes@caffertyclobes.com

JOSEPH SAVERI LAW FIRM, LLP
Joseph R. Saveri (SBN 130064)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com

DICELLO LEVITT LLP
Amy Keller (*pro hac vice*)
10 North Dearborn Street, Sixth Floor
Chicago, Illinois 60602
(312) 214-7900
akeller@dicellolevitt.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@butericklaw.com

*Counsel for Individual and Representative
 Plaintiffs and the Proposed Class*

(additional counsel included below)

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,
 	
   Individual and Representative Plaintiffs,
 	
v.
 	
META PLATFORMS, INC.,
 	
                        Defendant.
	Case No. 3:23-cv-03417-VC
	
	DECLARATION OF DAVID R. CHOFFNES, PH.D. IN SUPPORT OF PLAINTIFFS' ADMINISTRATIVE MOTION REQUESTING LEAVE TO OBTAIN LIMITED ADDITIONAL DISCOVERY

I, David R. Choffnes, Ph.D., declare as follows:

1.   I have personal knowledge of the matters stated herein and, if called upon, can competently testify thereto. I make this declaration pursuant to 28 U.S.C. § 1746 in support of Plaintiffs' Administrative Motion Requesting Leave to Obtain Limited Additional Discovery.

2.   I have been tasked with analyzing evidence produced by Meta with respect to torrenting of Plaintiffs' works between 2022 and 2024. During the course of my analysis, I have identified several instances of missing evidence that are needed to paint a complete picture of Meta's torrenting activity. It is my belief that Meta can produce additional evidence and admissions that will complete the record, thus allowing a firmer quantitative analysis of how much of Plaintiffs' works were reproduced and distributed during Meta's torrenting. The following paragraphs identify key areas where Meta has yet to produce sufficient evidence to make the record complete.

3.   **Additional 2022/2023 Torrenting Logs ("Lample Logs").** Meta produced tens of thousands of log files from Guillaume Lample's LibGen torrenting in 2022 and 2023. Many of these logs were irrelevant to torrenting, and represented other tasks for processing the book data.

4.   The logs that are relevant to torrenting indicate, for each second, how much of a torrent has been downloaded and how many bytes were uploaded and downloaded during that second. These logs are incomplete. Of the more than 70 TB of data I understand Lample to have torrented, the logs cover only 3.9 TB.

5.   Based on my analysis of these logs, there are 238 torrents where logs indicate a complete torrent download, but never show progress—i.e., the torrented data had been downloaded in a previous torrenting session for which Meta did not produce logs. This affects 31 copies of Plaintiffs' works. Without the logs, I cannot estimate how much of these works were distributed to other peers.

6.   In addition, Meta has produced only a limited set of logs related to LibGen Non-Fiction/SciTech torrenting in 2022.[1] However, contemporaneous evidence suggests that Lample torrented 70 TB of data, which would include LibGen Non-Fiction/SciTech.[2] There are 49 copies of asserted Plaintiffs' works in LibGen Non-Fiction/SciTech torrents where Meta has produced no logs.

7.   Due to the large scale of the logs produced (tens of thousands of documents), I was unable to identify the missing logs until recently.

8.   **Penguin Computing, AWS, and Azure Usage-based Billing Data.** If the missing log files are not produced, I can alternatively rely on information about how much data Meta uploaded and downloaded each day from its servers. This will allow me to identify when and where Lample's missing torrenting occurred, and subsequently estimate how much of Plaintiffs' works were distributed.

9.   Lample's code[3] indicates that he used Meta's H2 cluster, AWS, and Azure for his LibGen activities. For the H2 cluster, the King deposition[4] indicates that Penguin Computing was the provider. Meta has not produced usage logs from Penguin Computing. Meta has produced AWS usage logs for a subset of virtual servers in 2022, and the logs indicated torrenting was not occurring on those servers during that period. However, it is unclear if Lample used *different* AWS virtual servers, and additional AWS usage logs from the torrenting period are needed to determine this. Meta has not produced any usage logs from Azure.

---

[1] I understand Meta very recently discovered additional logs for 2024 torrenting of LibGen SciTech, which would be *in addition to* the 2022 torrenting of the same content.

[2] *See* Meta_Kadrey_00237829; *see also* the file at ▮▮▮▮▮▮▮▮ in the Lample code repository. Lample has several commands under "## Download" to torrent thousands of LibGen Non-Fiction/SciTech torrents in parallel. Lample's code repository also has a file at ▮▮▮▮▮▮▮▮ that processes SciTech books and refers to books being located in the "▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮" directory.

[3] Meta_Kadrey_00238129, in SciTech.md

[4] Christopher King April 2, 2025 Rule 30(b)(6) Dep. Tr. at 51:13-21.

10.   **Server Searches.** Lample's code makes references to several server directory locations where torrented books are stored. These include H2 directories "██" and ██", AWS directory "/████████████████████████████", and Azure directory ████████████████████████████████████ ████████████████████████████████". A listing of the files in these directories (and subdirectories), along with the file creation/modification dates, will allow me to identify when Plaintiffs' works were torrented, and on which computing platform. Along with usage logs, these listings of files torrented can be used to estimate how much of Plaintiffs' works were shared. It is my understanding that the directory listing, which Meta has already produced in other contexts, is straightforward to produce.

11.   It is important to note that Lample referred to LibGen using the term "BooksZero"[5] (alternatively "B0", "b0", "books0"), as seen in two of the file paths above. Between the fact that Lample's code was intentionally kept off of Meta's source code repository,[6] and the fact that he used code words like BooksZero to identify LibGen content, I did not identify these locations as relevant until recently.

12.   **RFAs.** I understand that Plaintiffs are requesting leave to serve additional RFAs to Meta. These RFAs would assist with my analysis of Meta's 2024 torrenting, specifically to confirm when, where, and how certain torrenting occurred given an incomplete record in prior production. Confirming the information in the RFAs is important to ensure the analysis of reproduction of

---

[5] I believe this to be a play on the names of the "Books1," Books2," and "Books3" collections of copyrighted work known to be used by other companies for AI training. See Darius Rafieyan and Hasan Chowdhury, *OpenAI destroyed a trove of books used to train AI models. The employees who collected the data are gone.*, Business Insider (May 7, 2024), https://www.businessinsider.com/openai-destroyed-ai-training-datasets-lawsuit-authors-books-copyright-2024-5.

[6] Meta_Kadrey_00237329 at -330: "I asked you if the Genesis [LibGen] data pipleines [*sic*] are anywhere in some repo [code repository] - so the pipelines are not anywhere because part of the pipelines touch the book torrents and we want that under wraps?"

Plaintiffs' works is correct. This is particularly important given the fact that Meta recently produced additional evidence regarding a new set of AWS instances where nearly 58 TB of torrenting occurred,[7] and very recently disclosed additional torrenting of LibGen Non-Fiction ("SciTech") in 2024. This new production and admission raised a number of new, important questions for analysis.

I declare under penalty of perjury that the foregoing is true and correct. Executed on the 29th day of October, 2025 in Needham, Massachusetts.

By: */s/ David R. Choffnes*
David R. Choffnes

---

[7] Meta_Kadrey_00238356 and Meta_Kadrey_00238357.

DECLARATION OF DAVID R. CHOFFNES
CASE NO. 3:23-cv-03417-VC